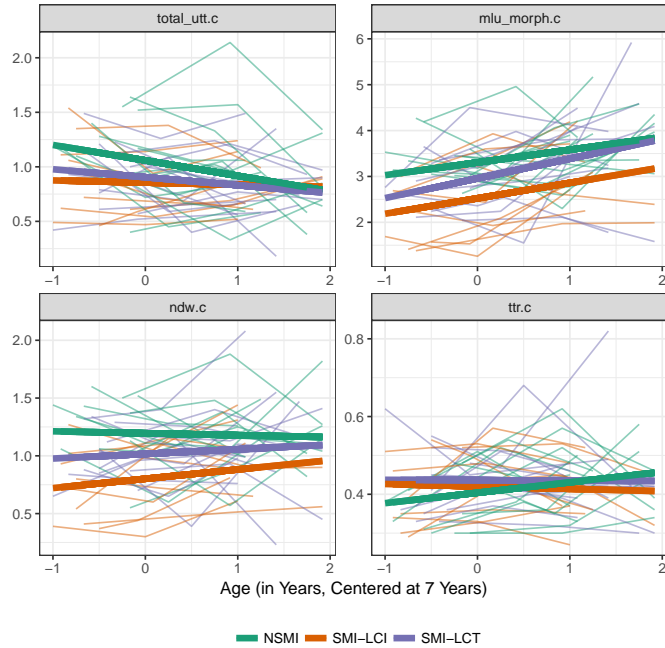


## Introduction

My name is Jacob Maronge. I am a 2nd year PhD Student in the Department of Statistics and Department of Biostatistics and Medical Informatics working with Professor Paul J. Rathouz. Generally, my research interests lie in the areas of experimental design and longitudinal or clustered data, including under situations with biased sampling schemes, such as outcome dependent sampling. Below I highlight a couple of the projects I am working on here at UW, one of an applied nature and one that is more methodological.

## Analysis of Speech Trajectories in Children with Cerebral Palsy

Paul and I collaborate with Professor Katie Hustad and her lab here at UW. A main project in Katie's lab involves following a longitudinal cohort of young children with Cerebral Palsy (CP) into adulthood while monitoring speech development compared to normally developing children. In some children, CP also affects their speech capabilities.

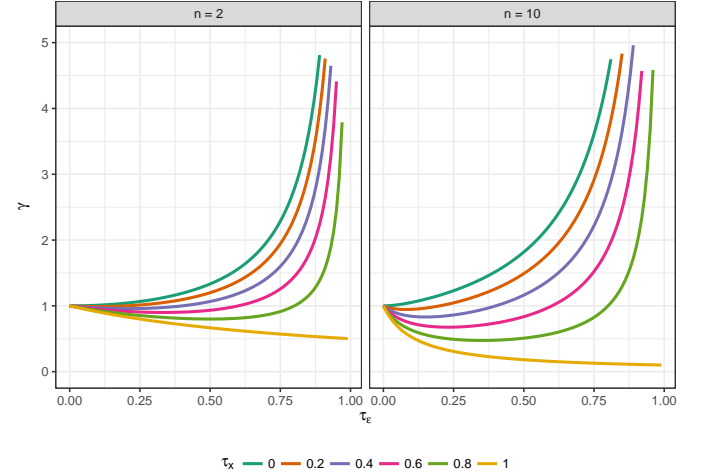


**Figure 1:** Each subject, color coded by group, for each outcome.

Clinicians categorize these children into 3 categories based off of speech abilities. One of our goals was to see if the data-driven methods regarding analysis of speech patterns collected on 35 children in the cohort suggested existence of these clinical categories. Figure 1 shows the subject level data for 4 speech variables of interest with the fitted values from a multivariate longitudinal model overlaid for each group.

## Power Analysis for Longitudinal Data Using Decomposition

Power analysis and sample size calculations are a crucial component of any study design. However, many study designs lead to either complicated formulas or do not have closed form solutions. We propose a method for calculating power in the situation of clustered or longitudinal data that leads to a solution that is both simple to calculate and gives valuable insight to how study design parameters affect power.



**Figure 2:** Parameter  $\gamma$  for different study design parameters,  $\tau_\epsilon$ ,  $\tau_x$ , and  $n$ .

We consider the situation where covariates are randomly observed. Our model of interest is,

$$Y_{i,j} = \beta_0 + \beta X_{i,j} + \epsilon_{i,j}, \quad (1)$$

where  $i = 1, \dots, m$  and  $j = 1, \dots, n$  denote the study participant and observation within participant, respectively. We expect measurements within subject to be positively correlated. Therefore,  $\text{corr}(X_{i,j}, X_{i,j'}) = \tau_x > 0$  and  $\text{corr}(\epsilon_{i,j}, \epsilon_{i,j'}) = \tau_\epsilon > 0$ . Finally,  $\text{Var}(X_{i,j}) = \text{Var}(\epsilon_{i,j}) = 1$ . Using an approach which decomposes the covariates, we are able to come to a power formula with two parts, one part is the correlation between  $X_{i,j}$  and  $Y_{i,j}$  without accounting for the correlation within subjects, and another part  $\gamma$ , is a function of study design parameters. This equation is shown below,

$$\rho_{\text{eff}}^2 = \frac{\rho^2 \gamma}{\rho^2 \gamma + (1 - \rho^2)}. \quad (2)$$

Figure 2 shows how  $\gamma$  varies as different study design parameters change. There are interesting special cases of this result. For example, when  $\tau_x = \tau_\epsilon$ , this is equivalent to an independent random sample, and when  $\tau_x = 1$  we recover the classical result for longitudinal studies comparing treatment to control group.